

-1-

Date: <u>Nov 29, 2001</u>	Express Mail Label No. <u>EV00536684245</u>
---------------------------	---

Inventor: Pedro J. Moreno

Attorney's Docket No.: 0918.2047-000

## SYSTEM AND METHOD FOR DETECTING REPETITIONS IN A MULTIMEDIA STREAM

### BACKGROUND OF THE INVENTION

Multimedia information streams such as streaming audio, video, and text are  
5 commonplace with the proliferation of information disseminated and available over  
information networks such as the Internet, telephone, cable TV, and wireless mediums.  
Massive amounts of multimedia data are transmitted over such networks, in the form of  
a digital stream, analog video, or text captioning, for example. Often, repetitions or  
near-repetitions of such data occur in these streams. Repetitions include transmissions  
10 such as paid advertisements, theme music at the commencement of a TV broadcast, and  
common jingles and slogans that may accompany transmissions from a common source.

Large amounts of multimedia data may be gathered by applications which store  
and process such data, such as SpeechBot™ and Mediaworqs™, for example.  
Repetitive transmissions can consume storage and computation resources redundantly if  
15 not detected. Also, processing of transmitted information, such as tracking paid  
advertisements to ensure frequency and duration, is typically performed by manually  
observing such multimedia transmissions. Detection and elimination or processing of  
repetitions can conserve resources, aid in tracking transmission patterns, and serve as  
building blocks for further processing. Accordingly, it would be beneficial to monitor  
20 and detect repetitions in a multimedia information stream to allow selective processing  
according to a specific application. One prior art technique for exact match audio

detection is disclosed in *Johnson, et al.*, "A Method for Direct Audio Search with Applications to Indexing and Retrieval," IEEE International Conference on Audio, Speech and Signal Processing (ICASSP 2000), June 5-9, 2000. Johnson, however, discloses a system which looks to a single vector derived from a portion of audio in  
5 relation to another single vector.

#### SUMMARY OF THE INVENTION

A method of detecting repetitions in an information stream of A/V (audio visual) data from a transmitted signal such as streaming audio or video includes extracting a plurality of samples from the information stream and accumulating the samples into  
10 segments comprising a predetermined interval of the transmitted signal. Vectors indicative of samples in respective segments are generated, and each of the vectors in the segments is correlated to generate a covariance matrix corresponding to the segment. The covariance matrices are aggregated into a sequence of covariance matrices and compared to each other covariance matrix in the sequence to generate a distance matrix.  
15 The distance matrix includes a distance value, indicative of the similarity between the covariance matrices, as a result of the comparing of each covariance matrix. The distance matrix is then traversed to determine similar sequences of covariance matrices, wherein determining similar sequences comprises searching for diagonals of similar distance values.

20 The distance matrix, therefore, contains a distance value for each pair of covariance matrices compared. A relatively low distance value between the two covariance matrices is indicative of a high degree of similarity. A sequence of relatively low distance values shows a repetition in the transmitted signal for an interval such as commercials, for example. Each segment represents a relatively small time interval so  
25 as to provide a high degree of granularity for the detection of repeated portions. Detection of repetitions, or duplicates, may include identification of near-duplicates which appear only slightly different due to sampling intervals or distortion. The higher granularity serves to ensure that the interval represented does not overlap with the start

or the end of the duplicate segment by only partial overlap. Graphically, such repetitions are represented as a diagonal in the distance matrix since they represent a contiguous sequence of time intervals corresponding to the compared matrices.

Found repetitions are stored in a library database for comparison with other  
5 information streams. The stream of samples is compared to itself and to the previously found repetitions in the library. A repository of candidates of likely repetitions is therefore maintained in the database. Matches may be either a new match or an instantiation of a previously found match. If a found sequence is a repeat of a  
10 library database is periodically scanned and entries older than a predetermined threshold are purged. The library is therefore limited to a manageable size by purging stale entries and refreshing current ones.

In this manner, repetitions of transmitted data may be detected and handled  
15 efficiently by ignoring, capturing, or otherwise processing the repetitions to conserve computing resources and avoid delays and interruptions resulting from undesired repetitions in the transmitted stream.

#### BRIEF DESCRIPTION OF THE DRAWINGS

The foregoing and other objects, features and advantages of the invention will be apparent from the following more particular description of preferred embodiments of  
20 the invention, as illustrated in the accompanying drawings in which like reference characters refer to the same parts throughout the different views. The drawings are not necessarily to scale, emphasis instead being placed upon illustrating the principles of the invention.

Fig. 1 is a context diagram of the repetition detection system as defined herein;  
25 Fig. 2 is a block diagram of the repetition detection system of the present invention;

Figs. 3a and 3b are schematic diagrams of covariance matrix processing;

Fig. 4 is a graphical illustration of a distance matrix;

Fig. 5 is an illustration of matching of sequences in a distance matrix;

Fig. 6 is an illustration of a distance matrix derived from a library; and

Figs. 7a and 7b are flowcharts of repetition detection processing of the present invention.

## 5 DETAILED DESCRIPTION OF THE INVENTION

A description of preferred embodiments of the invention follows.

A multimedia stream including audio, video, and text may contain repetitions of material such as advertisements and slogans. Identification of repeated material, or portions thereof, can allow the duplications to be removed or otherwise processed to  
10 avoid expending unnecessary computing resources to process the repeated material. Further, detection of repetitions may include identification of near-duplicates which appear only slightly different due to sampling intervals or distortion, and which otherwise represent the same information stream.

Fig. 1 shows a context diagram of the repetition detection system as defined  
15 herein. Referring to Fig. 1, the repetition detection system 10 receives a multimedia transmission stream 22 from a variety of sources, such as the Internet 12, magnetic or optical media 14, video recorder 16, or RF broadcasts from a receiver 18 via a PC 20. The transmission stream 22 may include video, audio, text, or a combination of these and other information carriers. The multimedia stream 22 is processed by the repetition  
20 detection system 10, which outputs an indication of the repetitions 24 for subsequent repetition processing. The repetition processing may include, for example, termination or skipping processing for duplicate portions, or recording frequency of occurrences of repetitions, for tracking purposes, loading a VCR compatible library module for advertisement detection, and the like.

25 Fig. 2 shows a block diagram of the detection repetition system 10. Referring to Fig. 2, the audio/visual (A/V) transmission stream 22 of multimedia data is received by a stream processor 30. The stream processor 30 subdivides the A/V stream 22 into samples 32 of the transmitted signal, each of a predetermined duration, or sampling

interval. In a particular embodiment, a sampling interval is 20  $\mu$ s, for example, but may be varied to suit a particular application. The samples 32 are sent to a segment processor 34, which accumulates samples 32 into segments representing an interval of transmission time. The segment processor 34 converts each sample 32 in the segment  
5 into a feature vector indicative of the transmitted signal over the sampling interval for each sample included in a segment. Each of the segments 36, therefore, comprises a set of vectors, or sequence vector set, corresponding to an interval of transmission time. In a particular embodiment, the transmission interval is 5 seconds, but may be varied according to the desired granularity, as will be described further below.

10 The segments 36 are sent to a correlator 38, which computes a covariance matrix, or signature 40, by correlating the vectors in the segment 36. Each of the signatures 40, therefore, tends to be uniquely indicative of the corresponding portion of the transmitted signal for the segment 36 (transmission interval). The signatures 40 are received by a distance processor 42, which determines signatures that are similar by  
15 comparing them to other signatures. Similarity is determined by computing a distance between signatures in the multidimensional space corresponding to the vectors. A library database 46 stores sequences which have previously found to be repeated, and which are therefore deemed to be likely candidates for further repetition, described further below. In a particular embodiment, the vectors correspond to a 39 dimensional  
20 space commonly employed for audio signals. More dimensions would typically be employed with a video signal, for example.

The distance processor 42 determines a measure of similarity between signatures 40, and generates distance matrices 44, described further below. The distance matrices 44 contain entries of the distance between signatures 40, or covariance matrices,  
25 generated from the transmission stream 22. A self distance matrix 50 and a library distance matrix 48 are generated. The self distance matrix 50 stores the distance values between signatures from the transmission stream 22 compared to itself. The library distance matrix 48 stores distance values between signatures 40 from the transmission stream 22 and signatures from known repetitive sequences stored in the library DB 46.

A traverser 54 receives the self distance matrix 50 and the library distance matrix 48, and identifies repetitions by searching for sequential entries of low distance values.

Once identified, new repetitions are stored in the library database 46 for use in successive distance matrices. A timestamp value ensures that stale entries are purged and frequently found sequences remain, so that the library 46 does not grow excessively large. The found duplicate sequences 56 are employed for successive processing applications, such as elimination of advertisements and tracking cycles of paid transmissions, for example. Also, new duplicates 52 are stored in the library 46 for use with successive transmission streams 22.

Figs. 3a and 3b show an example of covariance matrix processing (i.e., signature 40 generation) as disclosed in Fig. 2. Referring to Fig. 3a, the A/V stream 22 from the transmitted signal is subdivided into a plurality of segments 62a-62n, corresponding to time intervals  $t=1$  through  $t=T$ , respectively. Each of the segments 62a-62n includes a plurality of samples represented as a sequence vector set 64a-64n of respective vectors  $x_1 - x_D$ , each vector being generated from a sample of the transmission stream 22. In a particular embodiment each of the time intervals is 5 seconds of transmission time, and each of the samples is 1/100th of a second, hence there are 500 samples in a segment 62. Each of the vectors  $x_1 - x_D$  in the sequence vector sets 64a-64n is correlated with the other vectors  $x_1 - x_D$  in the sequence vector set 64a-64n to produce a covariance matrix 66a-66n, or signature, indicative of the respective sequence of vectors from the transmitted segment.

Fig. 3b shows covariance matrix processing in more detail employing sequence vector sets derived from overlapping and non-overlapping segments of samples.

Referring to Fig 3b, two alternate sequences of signatures are shown. A first sequence 66a'-66n' includes signatures 66 derived from non-overlapping sequence vector sets of samples, and corresponds to the signature sequence 66a-66n of Fig. 3a. The signatures 66a'-66n' are derived from non-overlapping sequence vector sets of samples, each sequence vector set 64a-64n including 500 contiguous samples from the transmission stream 22 as shown by the subscripts of the vectors  $x_1 - x_D$ , when  $D=500$ . Each

contiguous 500 samples comprises a sequence vector set 64a-64n. Specifically, signature 66a' is derived from samples  $x_1 - x_{500}$ , signature 66b' is derived from samples  $x_{501} - x_{1000}$ , and continuing to signature 66n', which is derived from samples  $x_{T-500} - x_T$ .

Continuing to refer to Fig. 3b, the sequence of signatures 66a''-66n'' includes  
 5 signatures derived from overlapping segments 62a-62n of samples. In this example, each segment includes 500 samples which overlap 250 samples with the adjacent sequence vector set 64. Therefore, signature 66a'' is derived from samples  $x_1 - x_{500}$ , however, signature 66b'' is derived from samples  $x_{250} - x_{750}$ , signature 66c'' is derived from samples  $x_{501} - x_{1000}$ , and continuing to signature 66n'', derived from samples  $x_{T-500} - x_T$ . Accordingly, the overlap processing derives more signatures but requires  
 10 additional storage.

Each covariance matrix 66a-66n can be represented as a vector in the multidimensional space, and may be compared to another vector to generate a distance value. As indicated above, the distance value is proportional to the similarity between  
 15 the transmitted segments 62a-62n, i.e., the closer in distance, and hence the lower in distance values, the greater the similarity between segments 62. Fig. 4 shows a distance matrix of distance values from a transmission stream 22 compared to itself. Referring to Fig. 4, the distance values are shown graphically according to a four-tier scale 71. A zero distance value between compared segments 62 is illustrated by squares on the scale  
 20 71. A low distance value between compared segments 62 is illustrated by hatch marks on the scale 71. A medium distance value is illustrated by diagonal lines, and a high distance value (i.e. a large dissimilarity) between compared segments 62 is illustrated by closed filled circle (dots) in the scale 71. A horizontal axis i 70 and a vertical axis j 72 each represent the same sequence of segments 62. Each graphed point ( i, j ) indicates  
 25 the distance between element i in the sequence to element j in the sequence. Since the streams being compared are the same, the upper half 76 is symmetrical with the lower half and is therefore not computed. A main diagonal, shown by dotted line 74, and illustrated by squares on the scale 71, has distance values of zero, which is typical when a stream is compared to itself.

Each element, or point (i, j), in the matrix, therefore, is compared to each other element to generate a distance value. The example shown employs four tiers of distance values 71, illustrated graphically, each corresponding to different distance threshold values. Varying numbers of tiers and thresholds may be employed. As indicated above, 5 in the example shown, the main diagonal has a distance value threshold of zero and generally is observed only when a segment is compared to itself. The low distance value tier has a threshold which allows distances that are sufficiently close to be considered a match, such as less than 0.5 or less than 1.0, depending on the application. Since the elements/matrix points (i, j) each represent sequential segments 62, a 10 matching sequence appears as a diagonal 78 of low distance values parallel to the main diagonal. Further, the diagonal 78 includes a sequence of a minimum number of segments corresponding to a minimum length of a repetition. In the example shown, the segments correspond to five seconds of transmission time, or transmission interval, therefore the diagonal 78 indicates a 15 second transmission. Other transmission 15 intervals corresponding to an expected duplicate transmission time could be employed, such as 20 or 30 seconds, described further below.

The low distance value tier threshold and the minimum number of repeated segments for a match define the granularity of the system. As described above, the product of the segment size (transmission interval) and the minimum number of 20 segments gives the minimum duration of a repeated segment found by the system. If the segment size is too large, the beginning of a repeated segment may not be detected until the following segment, and may be missed altogether if there is insufficient overlap. Similarly, if the minimum number of segments is too large, a shorter sequence of actual repetition may not be detected, possibly from only a portion of an advertisement having 25 been sampled. Conversely, if the minimum number of segments or the segment size are too small, repetitions may be indicated from trivial commonalities. In each case, the low distance tier threshold affects the degree of similarity, and therefore the likelihood of a near miss or trivial match, and may be employed to tune the sensitivity of the system accordingly.



The sequence of covariance matrices 66 corresponding to the transmitted stream may be compared to sequences of covariance matrices stored in the database 46 (Fig. 2) of previous transmissions, as well as to itself. Fig. 5 shows the matching of different sequences of covariance matrices. Referring to Fig. 5, a j axis 68 corresponds to increments of time of the transmitted stream 22 (Fig. 2). An i axis 92 corresponds to a sequence of covariance matrices from the library 46 of previous transmissions found to be repetitive. Since the covariance matrices from the transmitted stream are compared to a like number of covariance matrices representing a potential match, each axis of the compared segments represents the same number of segments, although not necessarily the same position in the sequence, as in the self distance matrix of Fig. 4.

In the example shown, segments 10-20 of the transmitted sequence 68 are compared to segments 70-80 of the library stream 92. This comparison is shown in region 94, in which a diagonal match 96 is found. The sequence of library segments shown by the i axis 92 may be of various sizes, depending on the number of previously found sequences and the available computing resources.

Fig. 6 shows an example of the invention repetition detection using a distance matrix of a transmitted stream and a library stream. Referring to Fig. 6, an i axis 80 corresponds to the library sequence, and a j axis 82 corresponds to the subject transmitted stream. Since the sequence of matrices is not being compared to itself, the axes are of unequal length and there is no main diagonal of zero distance, as illustrated above with respect to Fig. 4. A diagonal sequence 84 of low distance values illustrates a matching sequence. The sequence 84 includes eight (8) elements, indicating a repetitive transmission 40 seconds ( $8 * 5 \text{ seconds/element}$ ) in total duration. Another diagonal sequence 86 includes four elements, starting from the first element in the transmitted stream. Accordingly, 20 seconds are represented which may be only a partial sequence depending on the previous segments from the transmitted stream. Other matrix elements illustrating low distance values 88a-88c are either one or two elements in length. The minimum number of segments serves to remove apparent duplicates of a duration shorter than the non-trivial matches which are sought, such as the minimum

duration of an advertisement. Accordingly, matches of such short duration may not be indicative of a meaningful match and therefore, are not taken to indicate repetitions in the transmission stream.

Figs. 7a and 7b shows a flowchart of the duplicate detection routine of the present invention. Referring to Figs. 7a, 2 and 3, a new transmission stream 22 of A/V data is captured for duplication detection, as illustrated at step 100. Samples 32 of a predetermined sampling interval are taken from the stream 22, as shown at step 102. The samples 32 are accumulated into segments (Fig. 2, 36; Fig. 3a, 62), each segment 36 representing a minimal transmission interval, such as 5 seconds, as depicted at step 104. For each sample 32 in the segment 36, a feature vector ( $x_1 - x_D$  in Fig. 3a) is generated indicative of the sample 32, as disclosed at step 106, to produce a sequence vector set (64a-64n in Fig. 3a). As described above with respect to Figs. 3a and 3b, the feature vectors  $x_1 - x_D$  may correspond to either overlapping or non-overlapping sequence vector sets, depending on the indices of  $x_{(d)}$ . A check is performed to determine if any samples remain in the current segment 36, as shown at step 108. When all vectors ( $x_1 - x_D$  in Fig. 3a) have been determined for the segment 36, a covariance matrix (66a- 66n in Fig. 3a) is generated by correlating all the vectors in the sequence vector set (64a-64n) corresponding to the segment, as shown at step 110. The covariance matrix 66n is a signature 40 which tends to be uniquely indicative of the transmitted segment, and therefore is unlikely to yield a match against dissimilar transmissions. A check is performed to determine if more segments (36 Fig. 2, 62 Fig. 3a) remain in the transmitted stream 22, as disclosed at step 112, and control reverts to step 104, until the sequence of covariance matrices 66a-66n (Fig. 3a) corresponding to the transmitted stream 22 is completed. The sequence is then compared to the library 46 (Fig. 2) of sequences of previously found repetitions, and a library distance matrix generated, as depicted at step 114. The library distance matrix is generated first, before the self-distance matrix, so that found repetitions need not be searched again with respect to the self-distance matrix.

The library distance matrix is traversed to find diagonal sequences of low distance values, as depicted at step 115. A check is performed to see if a match is found in the library, as shown at step 116. If a match was found, then the matched sequence is marked in the transmitted stream 22, as disclosed at step 118, to avoid redundant searching in the self-match traversal. A time stamp in the library entry corresponding to the matched sequence is updated, as shown at step 120. As indicated above, the timestamp serves to keep the library entries from becoming stale. A predetermined timestamp threshold is employed to determine when entries are considered obsolete, such as one month. Other timestamp threshold values could be employed depending on the application. If there are more library sequences to compare, control reverts to step 115 to check the remaining sequences, as depicted at step 122.

Referring to Fig. 7b, after the library distance matrix has been traversed, the self distance matrix is computed by comparing the sequence of covariance matrices (generated through step 112) minus any matched sequence portions marked in step 118 to itself and generating the self distance matrix, as depicted at step 124. The self distance matrix is traversed to find diagonals of low distance values for at least the minimum threshold length, as shown at step 126. As indicated above, only the bottom half of this distance matrix is traversed since it is symmetrical across the main diagonal. A check is performed, to see if a matching sequence is found, as shown at step 128. If a match was found, the matching sequence is then written as a new entry to the library database of known sequences, as depicted at step 130. A timestamp is generated and stored with the entry, as disclosed at step 132. The timestamp is employed as described above with successive library matches. In an alternate embodiment, the self-distance matrix could be generated and traversed first, and the library database 46 traversed to find matches which are already known, in which case only the timestamp would be updated. A check is performed to see if the traversal is complete, as shown at step 134, otherwise control reverts to step 126 to find successive diagonals.

While this invention has been particularly shown and described with references to preferred embodiments thereof, it will be understood by those skilled in the art that

various changes in form and details may be made therein without departing from the scope of the invention encompassed by the appended claims. Accordingly, the invention is not intended to be limited except by the following claims.

Approved for Release by NSA on 08-26-2013 pursuant to E.O. 13526